BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PCT WORLD INTELLECTUAL PRO International

WO 9606110A1

INTERNATIONAL APPLICATION PUBLISHED UNDE

(51) International Patent Classification 6 : C07K 14/00, C12N 5/00, 15/00, C12P 21/06 — A1

(11) International Publication Number: WO 96/06110

(43) International Publication Date: 29 February 1996 (29.02.96)

(21) International Application Number: PCT/US95/10557

(22) International Filing Date: 18 August 1995 (18.08.95)

(30) Priority Data:

08/292,599	18 August 1994 (18.08.94)	US
08/373,351	17 January 1995 (17.01.95)	US
08/480,289	7 June 1995 (07.06.95)	US

(71) Applicant (*for all designated States except US*): ARIAD PHARMACEUTICALS, INC. [US/US]; 26 Landsdowne Street, Cambridge, MA 02139 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): GILMAN, Michael, Z. [CA/US]; 550 Chestnut Street, Newton, MA 02168 (US). NATESAN, Sridaran [IN/US]; 30 Thorton Road, Chestnut Hill, MA 02167 (US). POLLOCK, Roy, M. [GB/US]; Apartment 4C, 618 Boston Avenue, Medford, MA 02155 (US). BOTFIELD, Martyn, C. [US/US]; 44 Vernon Street, Brookline, MA 02146 (US).

(74) Agent: BERSTEIN, David; Ariad Pharmaceuticals, Inc., 26 Landsdowne Street, Cambridge, MA 02139 (US).

(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TT, UA, UG, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).

Published

With international search report.

(54) Title: COMPOSITE DNA-BINDING PROTEINS AND MATERIALS AND METHODS RELATING THERETO

(57) Abstract

This invention provides novel chimeric proteins and DNA sequences encoding them which are useful for regulated transcription of target genes in genetically engineered cells or organisms containing them. Target gene constructs and other materials useful for practicing the invention are also disclosed. Target gene constructs include a recombinant DNA sequence which is capable of binding to at least two heterologous DNA binding domains, e.g. in the form of a composite DNA binding protein or protein complex.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

Comp site DNA-Binding Pr teins and Materials and Methods Relating Thereto

Introduction

A large number of biological and clinical protocols, among others, gene therapy, production of biological materials, and biological research, depend on the ability to elicit specific and high-level expression of engineered genes encoding RNAs or proteins of therapeutic, commercial, or experimental value. This invention provides a method and materials for achieving high-level and controllable expression of such a target gene. The invention makes use of novel composite proteins containing multiple or composite DNA-binding domains designed to recognize, preferably with high affinity and specificity, DNA sequences associated with the target gene. This affinity and specificity are achieved by combining independent heterologous DNA-binding domains, by either covalent or non-covalent means, into a composite DNA-binding protein that recognizes a corresponding DNA sequence, preferably with very high affinity. By the further addition of a transcriptional activation domain to these proteins, by covalent or non-covalent means, the target gene can be activated to high levels of expression. At the same time, undesirable side-effects associated with the inadvertent activation of other genes is avoided.

Aspects of the design, production and use of biological switches based on ligand-mediated multimerization of immunophilin-based recombinant proteins are disclosed in Spencer et al, 1993, Science 262:1019-1024 and in PCT/US94/01617. This invention concerns new configurations for such biological switches and related methods and materials useful for regulated gene transcription, and applicable to constitutive gene expression as well. It involves recombinant DNA constructs, chimeric proteins encoded by the constructs, cells transformed with the constructs and methods for preparing and using the foregoing.

Chimeric proteins containing one or more ligand-binding domains atogether with DNA-binding or transcriptional activating domains are disclosed in PCT/US94/01617 and Spencer et al, supra. Those references provide substantial information, guidance and examples relating to the design, construction and use of DNA constructs encoding such chimeras, target gene constructs, multivalent ligands, and other aspects which may also be useful to

the practitioner of the subject invention. Their contents are incorporated herein by reference.

Summary of the invention

This disclosure focuses on the use of composite DNA-binding proteins, in which the component DNA-binding domains are covalently or non-covalently joined together, to obtain high level constitutive or regulated expression of a target gene for use in gene therapy, production of biological materials, and biological research. This invention involves novel DNA-binding proteins containing two or more heterologous DNA-binding domains which are linked together covalently or through an association mediated by a multimerizing agent (the terms "multimerize" and "dimerize" are used interchangably herein). The invention further involves DNA sequences encoding such proteins, the recombinant DNA sequences to which the composite DNA-binding proteins bind (i.e., which are recognized by the composite DNA-binding proteins), constructs containing a target gene and a DNA sequence which is recognized by the composite DNA-binding proteins, and the use of these materials in gene therapy, production of biological materials, and biological research. "Composite" as the term is used herein indicates that the protein contains component domains derived from at least two different proteins, domains from at least two non-adjacent portions of the same protein, or domains which are not found so linked in nature. Such composite proteins and DNA sequences which encode them are recombinant in the sense that they contain at least two constituent portions which are not otherwise found directly linked (covalently) together in nature. Desirable properties of these proteins include high affinity for specific DNA sequences, low affinity for most other sequences in a complex genome (such as human), low dissociation rates from specific DNA sites, and novel DNA recognition specificities distinct from those of known natural DNA-binding proteins. A basic principle of the design is the assembly of multiple DNA-binding domains into a single protein molecule or complex that recognizes a long and complex DNA sequence with high affinity through the combined interactions of the individual domains. A further benefit of this design is the avidity derived from multiple independent protein-DNA interactions. It bears repeating, and should be kept in mind by the reader, that the composite DNA binding protein in certain embodiments is a single chimeric protein containing multiple and

covalently-linked copies of one or more DNA-binding domains, while in other embodiments the composite DNA-binding protein comprises two (or more) "subunits", each of which is a chimeric protein in its own right containing at least one DNA-binding domain. In the latter case, the composite DNA-binding protein comprises two or more such subunits in a multimerizer-mediated association.

Components of the system

The system, as employed in cells, comprises: (1) a DNA construct encoding and directing the expression of a composite DNA-binding protein containing two or more heterologous component DNA-binding domains and one or more additional domains, as described below, or one or more DNA constructs encoding chimeric proteins, each containing one or more ligand-binding domains, DNA-binding domains and additional domains, which chimeras are capable of associating in the presence of a multimerizing agent, to form a composite DNA-binding protein (complex); (2) a DNA construct containing a target gene and one or more copies of a DNA sequence to which the composite DNA-binding protein is capable of binding, preferably with high affinity and/or specificity; and (3) optionally, one or more DNA constructs encoding and directing the expression of additional proteins capable of modulating the activity of the DNA-binding protein.

Preferably the composite DNA binding protein, whether formed by covalent linking or ligand-mediated multimerization of component parts, bind to a corresponding DNA sequence selectively, i.e., bind to that DNA sequence observable despite the presence of numerous alternative candidate DNA sequences. Preferably, binding of the multimerized chimeras or composite DNA binding protein to the selected DNA sequence is at least two, more preferably three and even more preferably more than four orders of magnitude greater than binding to any one alternative DNA sequence, as measured by relative rates or levels of transcription of genes associated with the selected and any alternative DNA sequences. It is also preferred that the selected DNA sequence be recognized to a substantially greater degree by the multimerized chimeras than by the non-multimerized chimeras, or by a protein containing a composite DBD than by a protein containing only some of the individual components thereof. Said differently, the level of expression of a target gene (in a cell containing the chimeric proteins and a target gene linked to a selected

DNA sequence) is preferably two, more preferably three and even more preferably more than four orders of magnitude greater in the presence of the multimerizing ligand than in its absence, as determined by any measure of transcription or target gene expression, including those described below. Likewise, target gene expression is preferably two, more preferably three, and even more preferably more than four orders of magnitude greater in the presence of a composite transcription factor containing a composite DBD than in the presence of a protein containing only some of the components of the composite DBD.

1. Design of Composite DNA-binding proteins.

(a) Covalently linked composite DBPs. Each covalently-linked or unitary composite DNA-binding protein consists of two or more protein domains capable of recognizing (i.e., binding to) specific sequences in DNA. The individual component domains may be separated by linker amino acids that permit the simultaneous contact of each domain with the DNA target. The combined action of the composite DNA-binding domain formed by the component DNA-binding modules results in the addition of the free energy decrement of each set of interactions. The effect is to achieve a DNA-protein interaction of very high affinity, preferably with dissociation constant below 10^{-9} M, more preferably below 10^{-10} M, even more preferably below 10^{-11} M. This goal is best achieved by combining domains that bind DNA poorly on their own, that is with low affinity, insufficient for functional recognition of DNA under typical conditions in a mammalian cell. Because the hybrid protein exhibits affinity for the composite site several orders of magnitude higher than the affinities of the individual sub-domains for their subsites, the protein preferentially (preferably exclusively) occupies composite sites.

Suitable component DNA-binding domains have one or more, preferably more, of the following properties. They bind DNA as monomers, although dimers can be accommodated. They should have modest affinities for DNA, with dissociation constants in the range of 10^{-6} to 10^{-9} M. They should optimally belong to a class of DNA-binding domains whose structure and interaction with DNA are well understood and therefore amenable to

manipulation. For gene therapy applications, they are preferably derived from human proteins.

(b) Multimerizer-linked composite DBPs. The multimerizer-linked composite DBPs comprise two or more chimeric proteins, each comprising at least one binding site for a multimerizing ligand, at least one component DBD such as mentioned above and described in further detail herein, and one or more optional domains, as discussed below. For background and additional practical details on the design, production and use of chimeric proteins containing ligand binding sites and capable of ligand-mediated multimerization, see e.g., Spencer et al, 1993, Science, supra, and PCT/US94/01617.

2. Examples of suitable component DNA-binding domains. DNA-binding domains with appropriate DNA binding properties may be selected from several different types of natural DNA-binding proteins. One class is proteins that normally bind DNA only in conjunction with auxiliary DNA-binding proteins, usually in a cooperative fashion, where both proteins contact DNA and each protein contacts the other. Examples of this class include the homeodomain proteins, many of which bind DNA with low affinity and poor specificity, but act with high levels of specificity in vivo due to interactions with partner DNA-binding proteins. One well-characterized example is the yeast alpha2 protein, which binds DNA only in cooperation with another yeast protein Mcm1. Another example is the human homeodomain protein Phox1, which interacts cooperatively with the human transcription factor, serum response factor (SRF).

A second class is proteins in which the DNA-binding domain is comprised of multiple reiterated modules that cooperate to achieve high-affinity binding of DNA. An example is the C_2H_2 class of zinc-finger proteins, which typically contain a tandem array of from two or three to dozens of zinc-finger modules. Each module contains an alpha-helix capable of contacting a three base-pair stretch of DNA. Typically, at least three zinc-fingers are required for high-affinity DNA binding. Therefore, one or two zinc-fingers constitute a low-affinity DNA-binding domain with suitable properties for use as a component in this invention. Examples of proteins of the C_2H_2 class include TFIIIA, Zif268, Gli, and SRE-ZBP. (These and other proteins and DNA

sequences referred to herein are well known in the art. Their sources and sequences are known.)

A third general class is proteins that themselves contain multiple independent DNA-binding domains. Often, any one of these domains is insufficient to mediate high-affinity DNA recognition, and cooperation with a covalently linked partner domain is required. Examples include the POU class, such as Oct-1, Oct-2 and Pit-1, which contain both a homeodomain and a POU-specific domain; HNF1, which is organized similarly to the POU proteins; certain Pax proteins (examples: Pax-3, Pax-6), which contain both a homeodomain and a paired box/domain; and XXX, which contains a homeodomain and multiple zinc-fingers of the C_2H_2 class.

An additional strategy for obtaining component DNA-binding domains with properties suitable for this invention is to modify an existing DNA-binding domain to reduce its affinity for DNA into the appropriate range. For example, a homeodomain such as that derived from the human transcription factor Phox1, may be modified by substitution of the glutamine residue at position 50 of the homeodomain. Substitutions at this position remove or change an important point of contact between the protein and one or two base pairs of the 6-bp DNA sequence recognized by the protein. Thus, such substitutions reduce the free energy of binding and the affinity of the interaction with this sequence and may or may not simultaneously increase the affinity for other sequences. Such a reduction in affinity is sufficient to effectively eliminate occupancy of the natural target site by this protein when produced at typical levels in mammalian cells. But it would allow this domain to contribute binding energy to and therefore cooperate with a second linked DNA-binding domain. Other domains that amenable to this type of manipulation include the paired box, the zinc-finger class represented by steroid hormone receptors, the myb domain, and the ets domain.

3. Design of linker sequence for covalently linked composite DBDs. The linker sequence separates adjacent DNA-binding domains. It should be selected or designed to permit the independent interaction of each domain with DNA without steric interference. A linker may also be selected or designed so as to impose specific spacing and orientation on the DNA-binding domains. The linker amino acids may be derived from endogenous flanking peptide sequence

of the component domains or may comprise one or more heterologous amino acids. Linkers may be designed by modeling or by experimental trial.

4. Additional domains. Additional domains may be included in the various chimeric proteins of this invention. For example, in some embodiments the chimeric proteins will contain a cellular targeting sequence which provides for the protein to be translocated to the nucleus. This nuclear localization sequence has a plurality of basic amino acids, referred to as a bipartite basic repeat (reviewed in Garcia-Bustos et al, Biochimica et Biophysica Acta (1991) 1071, 83-101). This sequence can appear in any portion of the molecule internal or proximal to the N- or C-terminus and results in the chimeric protein being inside the nucleus. The chimeric proteins may include domains that facilitate their purification, e.g. "histidine tags" or a glutathione-S-transferase domain. They may include "epitope tags" encoding peptides recognized by known monoclonal antibodies for the detection of proteins within cells or the capture of proteins by antibodies in vitro. They may also include one or more transcriptional activation domains, such as the well-characterized domain from the viral protein VP16 or novel activation domains of different designs. For instance, one may use one or multiple copies of transcriptional activating motifs from human proteins, including e.g. the 18 amino acid (NFLQLPQQTQGALLTSQP) glutamine rich region of Oct-2, the N-terminal 72 amino acids of p53, the SYGQQS repeat in Ewing sarcoma gene or an 11 amino acid (535-545) acidic rich region of Rel A protein. Chimeric proteins which contain both a composite DNA-binding domain and a transcriptional actibating domain thus comprise composite transcription factors.The chimeric proteins may include regulatory domains that place the function of the DNA-binding domain under the control of an external ligand; one example would be the ligand-binding domain of steroid receptors.

The chimeric proteins may also include a ligand-binding domain to provide for regulatable interaction of the protein with a second polypeptide chain. Thus, in embodiments involving covalently linked composite DNA binding domains, the unitary composite DNA-binding protein may further contain a ligand-binding domain. In such cases, the presence of a ligand-binding domain permits association of the composite DBP, in the presence of a dimerizing ligand, with a second chimeric protein containing a transcriptional activation domain and another ligand-binding domain. Alternatively, the

transcriptional activation domain may be present on a chimeric protein which further contains one or more component DNA-binding domains, which is capable of dimerizing, in the presence of a dimerizing agent, with another chimeric protein of this invention bearing a ligand-binding domain and one or more additional component DNA-binding domains. Upon dimerization of the chimeras a composite DNA-binding protein complex is formed which further contains the transcriptional activation domain and any other optional domains.

Multimerizing ligands useful in practicing this invention are multivalent, i.e., capable of binding to, and thus multimerizing, two or more of the chimeric protein molecules. The multimerizing ligand may bind to the chimeras containing such ligand-binding domains, in either order or simultaneously, preferably with a Kd value below about 10^{-6}, more preferably below about 10^{-7}, even more preferably below about 10^{-8}, and in some embodiments below about 10^{-9} M. The ligand preferably is not a protein or polypeptide and has a molecular weight of less than about 5 kDa, preferably below 2 kDa. The ligand-binding domains of the chimeric proteins so multimerized may be the same or different. See e.g. PCT/US93/01617, the full contents of which are hereby incorporated by reference.

5. Target DNA sequence. The DNA sequences recognized by the composite DNA-binding domains present in these proteins or protein complexes can be determined experimentally, as described below, or the proteins can be manipulated to direct their specificity toward a desired sequence. A desirable recognition sequence consists of at least twelve base pairs, preferably fifteen or even eighteen or more. These base pairs need not be fully contiguous; they may be interspersed with "spacer" base pairs that are not directly contacted by the protein but rather impose proper spacing between the subsites recognized by each module. These sequences should not impart expression to linked genes when introduced into cells in the absence of the engineered DNA-binding protein.

Design and assembly of the constructs

This section presents the general principles of design of system components and the details of the assembly of representative constructs of this invention.

1. General organization of composite DNA-binding d mains. The simplest organization for a functional composite DBD of this invention is two DNA-binding domains separated by a linker:

DBD1—L—DBD2

where each component domain (DBD) is independently capable of binding DNA with low affinity (dissociation constants in the range of 10^{-6} to 10^{-9} M), and the linker (L) is a stretch of amino acids of any length that permits a suitable orientation of the two DBDs on a single DNA molecule permitting binding to a target DNA with a dissociation constant below about 10^{-9}M.

For instance, one example comprises:

HD—L—ZF

where HD is a homeodomain (61 amino acids, with additional flanking sequences as necessary to obtain proper folding and stability) and ZF is one or two C_2H_2 zinc fingers separated by a natural zinc-finger linker (the H/C link). The boundaries of such domains are well characterized as is well known in the art.

Alternatively, these proteins can take the form of:

ZF—L—HD

where the domains are defined as above, but the order is different.

One currently preferred format is:

ZF1—L1—HD—L2—ZF2

In such cases three DNA-binding domains, as defined above, independently contact DNA. This arrangement is preferable, because it should result in higher affinities, slower dissociation rates, and larger recognition sequences.

Of course, for embodiments involving composite DNA-binding proteins formed only upon multimerizer-mediated assembly of a protein complex, each chimeric protein contains only a subset or portion of one of the foregoing

composite DBDs, together with other domains such as linker, ligand-binding, and other optional domains.

2. Modification of DNA-binding domains. Individual component DNA-binding domains may be further modified by mutagenesis to decrease, increase, or change the recognition specificity of DNA binding. These modifications could be achieved by rational design of substitutions in positions known to contribute to DNA recognition (often based on homology to related proteins for which explicit structural data are available). For example, in the case of a homeodomain, substitutions can be made in amino acids in the N-terminal arm, first loop, second helix, and third helix known to contact DNA. In zinc fingers, substitutions can be made at selected positions in the DNA recognition helix. Alternatively, random methods, such as selection from a phage display library could be used to identify altered domains with increased affinity or altered specificity.

3. Additional domains. Additional domains, described in the previous section (e.g., activation domains, ligand-binding domains) may be appended to either the N- or C-termini of the DNA-binding domains in any order consistent with the proper functioning of the protein (as may be readily observed experimentally).

4. Design and assembly of constructs. DNA sequences encoding individual DNA-binding sub-domains and linkers, if any, are joined such that they constitute a single open reading frame encoding a composite DBD that can be translated in cells or cell lysates into a single polypeptide harboring all domains. This protein-encoding sequence is then placed into a conventional plasmid vector that directs the expression of the protein in the appropriate cell type. For testing of proteins and determination of binding specificity and affinity, it may be desirable to construct plasmids that direct the expression of the protein in bacteria or in reticulocyte-lysate systems. For use in the production of proteins in mammalian cells, the protein-encoding sequence is introduced into an expression vector that directs expression in these cells. Expression vectors suitable for such uses are well known in the art. Various sorts of such vectors are commercially available.

In embodiments involving composite DNA-binding proteins formed by ligand-mediated multimerization rather than by covalent linkage, DNA sequences encoding a DNA-binding domain, with any introduced sequence alterations, is joined to DNA encoding one or more suitably engineered ligand-binding domains, and if desired, to DNA encoding a transcriptional activation domain or other optional domain(s). These sequences are joined such that they constitute a single open reading frame that can be translated in cells into a single polypeptide harboring all component domains. The order and arrangement of the domains within the polypeptide can vary. At least two such chimeras are required for the optimal embodiment of this method. These constructions encode polypeptides containing distinct DNA-binding domains, ligand-binding domains with distinct specificity for multimerizing moieties, and in some embodiments, transcriptional activation domains with different properties. For example, this invention includes chimeras of the following structure:

(immunophilin)---(txn activator)--(DNA binding domain)

wherein "immunophilin" represents 1, 2 or 3 immunophilin domains, such as the FKBP12 domain of Spencer et al, "txn activator" represents a VP16 domain and "DNA binding domain" represents a DNA binding domain of Phox1 or SRE-ZBP.

5. **Determination of target DNA sequences.** To identify a DNA sequence that is bound by the composite DNA-binding domain with high affinity (preferably with dissociation constant 10^{-11} M or lower), several methods can be used. If high-affinity binding sites for individual subdomains of the composite domain are already known, then these sequences can be joined with various spacing and orientation and the optimum configuration determined experimentally (see below for methods for determining affinities). Alternatively, high-affinity binding sites for the protein or protein complex can be selected from a large pool of random DNA sequences by adaptation of published methods (Pollock, R. and Treisman, R., 1990, A sensitive method for the determination of protein-DNA binding specificities. Nucl. Acids Res. 18, 6197-6204). Bound sequences are cloned into a plasmid and their precise sequence and affinity for the proteins are determined. From this collection of

sequences, individual sequences with desirable characteristics (i.e., maximal affinity for composite protein, minimal affinity for individual subdomains) are selected for use. Alternatively, the collection of sequences is used to derive a consensus sequence that carries the favored base pairs at each position. Such a consensus sequence is synthesized and tested (see below) to confirm that it has an appropriate level of affinity and specificity.

6. Design of target gene. The DNA construct that enables the target gene to be regulated by DNA-binding proteins of this invention is a fragment, plasmid, or other nucleic acid vector carrying a synthetic transcription unit consisting of: (1) one copy or multiple copies of a DNA sequence recognized with high-affinity by the composite DNA-binding protein or protein complex; (2) a promoter sequence consisting minimally of a TATA box and initiator sequence but optionally including other transcription factor binding sites; (3) sequence encoding the desired product (protein or RNA), including sequences that promote the initiation and termination of translation, if appropriate; (4) an optional sequence consisting of a splice donor, splice acceptor, and intervening intron DNA; and (5) a sequence directing cleavage and polyadenylation of the resulting RNA transcript.

Testing and optimization

This section describes methods for evaluating the efficacy of DNA-binding proteins designed according to the principles of this invention and strategies for optimization of the protein-DNA interaction.

1. Determination of binding affinity. A number of well-characterized assays are available for determining the binding affinity, usually expressed as dissociation constant, for DNA-binding proteins and their cognate DNA sequences. These assays usually require the preparation of purified protein and binding site (usually a synthetic oligonucleotide) of known concentration and specific activity. Examples include electrophoretic mobility-shift assays, DNaseI protection or "footprinting", and filter-binding. These assays can also be used to get rough estimates of association and dissociation rate constants. These values may be determined with greater precision using a BIAcore instrument. In this assay, the synthetic oligonucleotide is bound to the assay "chip," and purified DNA-binding protein is passed through the flow-cell. Binding of the protein to

the DNA immobilized on the chip is measured as an increase in refractive index. Once protein is bound at equilibrium, buffer without protein is passed over the chip, and the dissociation of the protein results in a return of the refractive index to baseline value. The rates of association and dissociation are calculated from these curves, and the affinity or dissociation constant is calculated from these rates. Binding rates and affinities for the high affinity composite site may be compared with the values obtained for subsites recognized by each subdomain of the protein. As noted above, the difference in these dissociation constants should be at least two orders of magnitude and preferably three or greater.

2. Testing for function in vivo. Several tests of increasing stringency may be used to confirm the satisfactory performance of a DNA-binding protein (or complex) designed according to this invention. All share essentially the same components: (1) (a) an expression plasmid directing the production of a fusion protein composed of the novel composite DNA-binding domain and a potent transcriptional activation domain (i.e., a composite transcription factor) or (b) one or more expression plasmids directing the production of a pair of chimeric proteins of this invention which are capable of dimerizing in the presence of a corresponding dimerizing agent, and thus forming a composite DNA-binding protein complex; and (2) an expression plasmid directing the expression of a reporter gene, preferably identical in design to the target gene described above (i.e., multiple binding sites for the DNA-binding domain, a minimal promoter element, and a gene body) but encoding any conveniently measured protein.

In a transient transfection assay, the above-mentioned plasmids are introduced together into tissue culture cells by any conventional transfection procedure, including for example calcium phosphate coprecipitation, electroporation, and lipofection. After an appropriate time period, usually 24–48 hr, the cells are harvested and assayed for production of the reporter protein. In embodiments requiring dimerization of chimeric proteins for activation of transcription, the assay is conducted in the presence of the dimerizing agent. In an appropriately designed system, the reporter gene should exhibit little activity above background in the absence of any co-transfected plasmid for the composite transcription factor (or in the absence of dimerizing agent in embodiments under dimerizer control). In contrast, reporter gene expression should be elevated in a dose-dependent fashion by the inclusion of the plasmid

encoding the composite transcription factor (or plasmids encoding the multimerizable chimeras, followed by multimerizing agent). This result indicates that there are few natural transcription factors in the recipient cell with the potential to recognize the tested binding site and activate transcription and that the engineered DNA-binding domain is capable of binding to this site inside living cells.

The transient transfection assay is not a stringent test in most cases, because the high concentrations of plasmid DNA in the transfected cells lead to unusually high concentrations of the DNA-binding protein and its recognition site, allowing functional recognition even with relative low affinity interactions. A more stringent test of the system is a transfection that results in the integration of the introduced DNAs at near single-copy. Thus, both the protein concentration and the ratio of specific to non-specific DNA sites would be very low; only very high affinity interactions would be expected to be productive. This scenario is most readily achieved by stable transfection in which the plasmids are transfected together with another plasmid encoding an unrelated selectable marker (e.g., G418-resistance). Transfected cell clones selected for drug resistance typically contain copy numbers of the nonselected plasmids ranging from zero to a few dozen. A set of clones covering that range of copy numbers can be used to obtain a reasonably clear estimate of the efficiency of the system.

Perhaps the most stringent test involves the use of a viral vector, typically a retrovirus, that incorporates both the reporter gene and the gene encoding the composite transcription factor or multimerizable components thereof. Virus stocks derived from such a construction will generally lead to single-copy transduction of the genes.

If the ultimate application is gene therapy, it may be preferred to construct transgenic animals carrying similar DNAs to determine whether the protein is functional in an animal.

3. Optimization. Once a composite DNA-binding domain (covalently linked or formed by dimerization) and a corresponding DNA sequence are obtaining, further engineering of the DNA binding protein or its components may be done with the goal of increasing the affinity or changing the sequence specificity of the interaction. One simple step is the addition of a further DNA-binding module. A single zinc-finger, for example, will typically add three more

base pairs to the DNA recognition sequence and raise the affinity of the interaction. Inspection of the structure of related domains may suggest amino acid substitutions that increase binding affinity. Alternatively, a random strategy, such as selection from a phage display library, may be used to select high-affinity protein variants. This strategy is particularly effective with zinc fingers.

In addition, the recognition specificity of the protein can be changed. Substituting the amino acid at position 50 of a homeodomain, changes the recognition specificity for positions 5 and 6 in the 6 base-pair binding site. Similar mutations in the recognition helix of zinc fingers also change DNA recognition specificity. In the case of zinc fingers, phage display has been used effectively to select zinc fingers that recognize a given three base-pair sequence.

Introduction of Constructs into Cells

Constructs encoding the composite DNA-binding proteins, constructs encoding related chimeric proteins (e.g. in the case of regulatable expression systems) and constructs directing the expression of target genes, all as described herein, can be introduced into cells as one or more DNA molecules or constructs, in many cases in association with one or more markers to allow for selection of host cells which contain the construct(s). The constructs can be prepared in conventional ways, where the coding sequences and regulatory regions may be isolated, as appropriate, ligated, cloned in an appropriate cloning host, analyzed by restriction or sequencing, or other convenient means. Particularly, using PCR, individual fragments including all or portions of a functional unit may be isolated, where one or more mutations may be introduced using "primer repair", ligation, *in vitro* mutagenesis, *etc.* as appropriate. The construct(s) once completed and demonstrated to have the appropriate sequences may then be introduced into a host cell by any convenient means. The constructs may be integrated and packaged into non-replicating, defective viral genomes like Adenovirus, Adeno-associated virus (AAV), or Herpes simplex virus (HSV) or others, including retroviral vectors, for infection or transduction into cells. The constructs may include viral sequences for transfection, if desired. Alternatively, the construct may be introduced by fusion, electroporation, biolistics, transfection, lipofection, or the like. The host cells will in some cases be grown and expanded in culture before introduction of the construct(s), followed by the appropriate treatment for

introduction of the construct(s) and integration of the construct(s). The cells will then be expanded and screened by virtue of a marker present in the construct. Various markers which may be used successfully include *hprt*, neomycin resistance, thymidine kinase, hygromycin resistance, *etc.*

In some instances, one may have a target site for homologous recombination, where it is desired that a construct be integrated at a particular locus. For example, one can delete and/or replace an endogenous gene (at the same locus or elsewhere) with a recombinant target construct of this invention. For homologous recombination, one may generally use either Ω or O-vectors. See, for example, Thomas and Capecchi, *Cell* (1987) 51, 503-512; Mansour, *et al.*, *Nature* (1988) 336, 348-352; and Joyner, *et al.*, *Nature* (1989) 338, 153-156.

The constructs may be introduced as a single DNA molecule encoding all of the genes, or different DNA molecules having one or more genes. The constructs may be introduced simultaneously or consecutively, each with the same or different markers.

Vectors containing useful elements such as bacterial or yeast origins of replication, selectable and/or amplifiable markers, promoter/enhancer elements for expression in procaryotes or eucaryotes, etc. which may be used to prepare stocks of construct DNAs and for carrying out transfections are well known in the art, and many are commercially available.

Introduction of Constructs into Animals

Cells which have been modified ex vivo with the DNA constructs may be grown in culture under selective conditions and cells which are selected as having the desired construct(s) may then be expanded and further analyzed, using, for example, the polymerase chain reaction for determining the presence of the construct in the host cells. Once modified host cells have been identified, they may then be used as planned, e.g. grown in culture or introduced into a host organism.

Depending upon the nature of the cells, the cells may be introduced into a host organism, e.g. a mammal, in a wide variety of ways. Hematopoietic cells may be administered by injection into the vascular system, there being usually at least about 10^4 cells and generally not more than about 10^{10}, more usually not more than about 10^8 cells. The number of cells which are employed will depend upon a number of circumstances, the purpose for the introduction, the lifetime of the cells, the protocol to be used, for example, the number of

administrations, the ability of the cells to multiply, the stability of the therapeutic agent, the physiologic need for the therapeutic agent, and the like. Alternatively, with skin cells which may be used as a graft, the number of cells would depend upon the size of the layer to be applied to the burn or other lesion. Generally, for myoblasts or fibroblasts, the number of cells will be at least about 10^4 and not more than about 10^8 and may be applied as a dispersion, generally being injected at or near the site of interest. The cells will usually be in a physiologically-acceptable medium.

Cells engineered in accordance with this invention may also be encapsulated, e.g. using conventional materials and methods. See e.g. Uludag and Sefton, 1993, J Biomed. Mater. Res. 27(10):1213-24; Chang et al, 1993, Hum Gene Ther 4(4):433-40; Reddy et al, 1993, J Infect Dis 168(4):1082-3; Tai and Sun, 1993, FASEB J 7(11):1061-9; Emerich et al, 1993, Exp Neurol 122(1):37-47; Sagen et al, 1993, J Neurosci 13(6):2415-23; Aebischer et al, 1994, Exp Neurol 126(2):151-8; Savelkoul et al, 1994, J Immunol Methods 170(2):185-96; Winn et al, 1994, PNAS USA 91(6):2324-8; Emerich et al, 1994, Prog Neuropsychopharmacol Biol Psychiatry 18(5):935-46 and Kordower et al, 1994, PNAS USA 91(23):10898-902. The cells may then be introduced in encapsulated form into an animal host, preferably a mammal and more preferably a human subject in need thereof. Preferably the encapsulating material is semipermeable, permitting release into the host of secreted proteins produced by the encapsulated cells. In many embodiments the semipermeable encapsulation renders the encapsulated cells immunologically isolated from the host organism in which the encapsulated cells are introduced. In those embodiments the cells to be encapsulated may express one or more chimeric proteins containing components domains derived from viral proteins or proteins from other species (and need not contain a composite DNA binding domain as described above). For example in those cases the chimeras may contain elements derived from GAL4 and VP16. In such cases, the cells may be engineered as disclosed in International Patent Applications PCT/US94/01617 or PCT/US94/08008 or in US Patent Application Serial Nos. 08/292,595 and 08/292,596 (filed August 18, 1994), the full contents of which are incorporated herein by reference.

Instead of *ex vivo* modification of the cells, in many situations one may wish to modify cells *in vivo*. For this purpose, various techniques have been developed for modification of target tissue and cells *in vivo*. A number of virus vectors have been developed, such as adenovirus, adeno-associated virus.

and retroviruses, which allow for transfection and random integration of the virus into the host. See, for example, Debunks et al. (1984) Porch. Natl. Acad. Sci. USA 81, 7529-7533; Caned et al., (1989) Science 243,375-378; Hiebert et al. (1989) Proc. Natl. Acad. Sci. USA 86, 3594-3598; Hatzoglu et al. (1990) J. Biol. Chem. 265, 17285-17293 and Ferry, et al. (1991) Proc. Natl. Acad. Sci. USA 88, 8377-8381. The vector may be administered by injection, e.g. intravascularly or intramuscularly, inhalation, or other parenteral mode.

In accordance with *in vivo* genetic modification, the manner of the modification will depend on the nature of the tissue, the efficiency of cellular modification required, the number of opportunities to modify the particular cells, the accessibility of the tissue to the DNA composition to be introduced, and the like. By employing an attenuated or modified retrovirus carrying a target transcriptional initiation region, if desired, one can activate the virus using one of the subject transcription factor constructs, so that the virus may be produced and transfect adjacent cells.

The DNA introduction need not result in integration in every case. In some situations, transient maintenance of the DNA introduced may be sufficient. In this way, one could have a short term effect, where cells could be introduced into the host and then turned on after a predetermined time, for example, after the cells have been able to home to a particular site.

Applications

This invention is applicable to any situation that calls for expression of an exogenously-introduced gene embedded within a large genome. The desired expression level could be preset very high or very low. Alternatively, the system may be further engineered to achieve regulated or titratable expression. See e.g. PCT/US93/01617. In most cases, the inadvertent activation of unrelated cellular genes is undesirable.

1. Constitutive high-level gene expression in gene therapy. Gene therapy often requires controlled high-level expression of a therapeutic gene, sometimes in a cell-type specific pattern. By supplying the therapeutic gene with saturating amounts of an activating transcription factor of this invention, considerably higher levels of gene expression can be obtained relative to natural promoters or enhancers, which are dependent on endogenous transcription factors. Thus, one application of this invention to gene therapy is the delivery

of a two-transcription-unit cassette (which may reside on one or two plasmid molecules, depending on the delivery vector) consisting of (1) a transcription unit encoding a protein composed of a composite DNA-binding domain constructed according to this invention and a strong transcription activation domain (e.g., derived from the VP16 protein) and (2) a transcription unit consisting of the therapeutic gene expressed under the control of a minimal promoter carrying one, and preferably several, binding sites for the composite DNA-binding domain. Cointroduction of the two transcription units into a cell results in the production of the hybrid transcription factor which in turn activates the therapeutic gene to high level. This strategy essentially incorporates an amplification step, because the promoter that would be used to produce the therapeutic gene product in conventional gene therapy is used instead to produce the activating transcription factor. Each transcription factor has the potential to direct the production of multiple copies of the therapeutic protein.

This method may be employed to increase the efficacy of many gene therapy strategies by substantially elevating the expression of the therapeutic gene, allowing expression to reach therapeutically effective levels. Examples of therapeutic genes that would benefit from this strategy are genes that encode secreted therapeutic proteins, such as cytokines (e.g., IL-2, IL-4, IL-12), growth factors (e.g., VEGF), antibodies, and soluble receptors. Other candidate therapeutic genes are disclosed in PCT/US93/01617. This strategy may also be used to increase the efficacy of "intracellular immunization" agents, molecules like ribozymes, antisense RNA, and dominant-negative proteins, that act either stoichiometrically or by competition. Examples include agents that block infection by or production of HIV or hepatitis virus and agents that antagonize the production of oncogenic proteins in tumors.

2. Regulated gene therapy. In many instances, the ability to switch a therapeutic gene on and off at will or the ability to titrate expression with precision are absolutely essential to therapeutic efficacy. This invention is particularly well suited for achieving regulated expression of a target gene. Two examples of how regulated expression may be achieved are described. The first involves a recombinant transcription factor which comprises a composite DNA-binding domain, a potent transcriptional activation domain, and a regulatory domain controllable by a small orally-available ligand. One example

is the ligand-binding domain of steroid receptors, in particular the domain derived from the modified progesterone receptor described by Wang et al, 1994, Proc Natl Acad Sci USA 91:8180-8184. In this example, the composite DNA binding domain of this invention is used in place of the GAL4 domain in the recombinant transcription factor and the target gene is linked to a DNA sequence recognized by the composite DNA binding domain. Such a design permits the regulation of a target gene by known anti-progestins such as RU486. The transcription factors described here greatly enhance the efficacy of this regulatory domain because of the enhanced affinity of the DNA-binding domain, the absence of background activity that arises from ligand-independent dimerization directed by the GAL4 domain in published constructs, and the reduced potential for immunogenicity because human sequences are substituted for yeast.

Another example involves a pair of chimeric proteins, a dimerizing agent capable of dimerizing the chimeras and a target gene construct to be expressed. The first chimeric protein comprises a composite DNA-binding domain as described herein and a receptor domain (e.g. FKBP) for which a ligand, preferably a high-affinity ligand, is available. The second chimeric protein comprises an activation domain and a second receptor domain (which may be the same or different than on the prior chimeric protein). The dimerizing reagent is capable of binding to the receptor (or "ligand binding") domains present on each of the chimeras and thus of dimerizing or oligomerizing the chimeras. DNA molecules encoding and directing the expression of these chimeric proteins are introduced into the cells to be engineered. Also introduced into the cells is a target gene linked to a DNA sequence to which the composite DNA-binding domain is capable of binding. Contacting the engineered cells or their progeny with the oligomerizing reagent leads to regulated activity of the transcription factor and hence to expression of the target gene. The design and use of similar components is disclosed in PCT/US93/01617. These may be adapted to the present invention by the use of a composite DNA-binding domain, and DNA sequence encoding it, in place of the alternative DNA-binding domains as disclosed in the referenced patent document.

The dimerizing ligand may be administered to the patient as desired to activate transcription of the target gene. Depending upon the binding affinity of the ligand, the response desired, the manner of administration, the half-life,

the number of cells present, various protocols may be employed. The ligand may be administered parenterally or orally. The number of administrations will depend upon the factors described above. The ligand may be taken orally as a pill, powder, or dispersion; bucally; sublingually; injected intravascularly, intraperitoneally, subcutaneously; by inhalation, or the like. The ligand (and monomeric antagonist compound) may be formulated using conventional methods and materials well known in the art for the various routes of administration. The precise dose and particular method of administration will depend upon the above factors and be determined by the attending physician or human or animal healthcare provider. For the most part, the manner of administration will be determined empirically.

In the event that transcriptional activation by the ligand is to be reversed or terminated, a monomeric compound which can compete with the dimerizing ligand may be administered. Thus, in the case of an adverse reaction or the desire to terminate the therapeutic effect, an antagonist to the dimerizing agent can be administered in any convenient way, particularly intravascularly, if a rapid reversal is desired. Alternatively, one may provide for the presence of an inactivation domain (or transcriptional silencer) with a DNA binding domain. In another approach, cells may be eliminated through apoptosis via signaling through Fas or TNF receptor as described elsewhere. See International Patent Applications PCT/US94/01617 and PCT/US94/08008.

The particular dosage of the ligand for any application may be determined in accordance with the procedures used for therapeutic dosage monitoring, where maintenance of a particular level of expression is desired over an extended period of times, for example, greater than about two weeks, or where there is repetitive therapy, with individual or repeated doses of ligand over short periods of time, with extended intervals, for example, two weeks or more. A dose of the ligand within a predetermined range would be given and monitored for response, so as to obtain a time-expression level relationship, as well as observing therapeutic response. Depending on the levels observed during the time period and the therapeutic response, one could provide a larger or smaller dose the next time, following the response. This process would be iteratively repeated until one obtained a dosage within the therapeutic range. Where the ligand is chronically administered, once the maintenance dosage of the ligand is determined, one could then do assays at extended intervals to be

assured that the cellular system is providing the appropriate response and level of the expression product.

It should be appreciated that the system is subject to many variables, such as the cellular response to the ligand, the efficiency of expression and, as appropriate, the level of secretion, the activity of the expression product, the particular need of the patient, which may vary with time and circumstances, the rate of loss of the cellular activity as a result of loss of cells or expression activity of individual cells, and the like. Therefore, it is expected that for each individual patient, even if there were universal cells which could be administered to the population at large, each patient would be monitored for the proper dosage for the individual.

3. Production of recombinant proteins and viruses. Production of recombinant therapeutic proteins for commercial and investigational purposes is often achieved through the use of mammalian cell lines engineered to express the protein at high level. The use of mammalian cells, rather than bacteria or yeast, is indicated where the proper function of the protein requires post-translational modifications not generally performed by heterologous cells. Examples of proteins produced commercially this way include erythropoietin, tissue plasminogen activator, clotting factors such as Factor VIII:c, antibodies, etc. The cost of producing proteins in this fashion is directly related to the level of expression achieved in the engineered cells. Thus, because the constitutive two-transcription-unit system described above can achieve considerably higher expression levels than conventional expression systems, it may greatly reduce the cost of protein production. A second limitation on the production of such proteins is toxicity to the host cell: Protein expression may prevent cells from growing to high density, sharply reducing production levels. Therefore, the ability to tightly control protein expression, as described for regulated gene therapy, permits cells to be grown to high density in the absence of protein production. Only after an optimum cell density is reached, is expression of the gene activated and the protein product subsequently harvested.

A similar problem is encountered in the construction and use of "packaging lines" for the production of recombinant viruses for commercial (e.g., gene therapy) and experimental use. These cell lines are engineered to produce viral proteins required for the assembly of infectious viral particles harboring defective recombinant genomes. Viral vectors that are dependent on

such packaging lines include retrovirus, adenovirus, and adeno-associated virus. In the latter case, the titer of the virus stock obtained from a packaging line is directly related to the level of production of the viral rep and core proteins. But these proteins are highly toxic to the host cells. Therefore, it has proven difficult to generate high-titer recombinant viruses. This invention provides a solution to this problem, by allowing the construction of packaging lines in which the rep and core genes are placed under the control of regulatable transcription factors of the design described here. The packaging cell line can be grown to high density, infected with helper virus, and transfected with the recombinant viral genome. Then, expression of the viral proteins encoded by the packaging cells is induced by the addition of dimerizing agent to allow the production of virus at high titer.

4. Biological research. This invention is applicable to a wide range of biological experiments in which precise control over a target gene is desired. These include: (1) expression of a protein or RNA of interest for biochemical purification; (2) regulated expression of a protein or RNA of interest in tissue culture cells for the purposes of evaluating its biological function; (3) regulated expression of a protein or RNA of interest in transgenic animals for the purposes of evaluating its biological function; (4) regulating the expression of another regulatory protein that acts on an endogenous gene for the purposes of evaluating the biological function of that gene. Transgenic animal models and other applications in which the composite DNA-binding domains of this invention may be used include those disclosed in US Patent Application Serial Nos. 08/292,595 and 08/292,596 (filed August 18, 1994).

This invention further provides kits useful for the foregoing applications. Such kits contain a first DNA sequence encoding a recombinant protein comprising a composite DNA binding domain of this invention (and may contain additional domains as discussed above) and a second DNA sequence containing a target gene linked to a DNA element to which the recombinant protein is capable of binding. Alternatively, the second DNA sequence may contain a cloning site for insertion of a desired target gene by the practitioner. For regulatable applications, i.e., in cases in which the recombinant protein contains a composite DNA-binding domain and a receptor domain, the kit may further contain a third DNA sequence encoding a transcriptional

activating domain and a second receptor domain, as discussed above. Such kits may also contain a sample of a dimerizing agent capable of dimerizing the two recombinant proteins and activating transcription of the target gene.

The following examples contain important additional information, exemplification and guidance which can be adapted to the practice of this invention in its various embodiments and the equivalents thereof. The examples are offered by way illustration and not by way limitation.

Examples

The following examples detail the construction of DNA vectors containing recombinant DNA sequences encoding component DNA binding subdomains and composite DNA binding domains of this invention. The constructs encoding the composite DNA binding domains may be linked to other elements and used in the various applications disclosed herein.

Constructs.

All plasmids are constructed in pET-19BHA, a pET-19B based vector modified such that all expressed proteins contain an amino-terminal Histidine "Tag" for purification and an epitope tag for immunoprecipitation. pET-19B is a well-known vector for expression of heterologous proteins in E coli or in reticulocyte lysates.

Zinc Finger Constructs

All zinc finger sequences are derived from the human cDNA encoding SRE-ZBP (Attar, R.M. and Gilman, M.Z. 1992. MCB 12: 2432-2443).

p19B2F: Contains SREZBP zinc fingers 6 and 7 (amino acids 328 to 410) fused in frame to the epitope tag in p19BHA. DNA encoding ZBP zinc fingers 6 and 7 was generated by PCR using primers 2F-Xba5' and ZNF-Spe/Bam (see below). The resulting fragment was cut with XbaI and BamHI and ligated between the XbaI and BamHI sites of pET-19BHA.

p19B4F: Contains SREZBP zinc fingers 4, 5, 6 and 7 (amino acids 300 to 410) fused in frame to the epitope tag in p19BHA. A DNA fragment encoding ZBP zinc fingers 4, 5, 6 and 7 was generated by PCR using primers 4F-Xba5' and ZNF-

Spe/Bam. The resulting fragment was cut with XbaI and BamHI and ligated between the XbaI and BamHI sites of pET-19BHA.

p19B7F: Contains SREZBP zinc fingers 1 to 7 (amino acids 216 to 410) fused in frame to the epitope tag in p19BHA. DNA encoding ZBP zinc fingers 1 to 7 was generated by PCR using primers 7F-Xba5' and ZNF-Spe/Bam. The resulting fragment was cut with XbaI and BamHI and ligated between the XbaI and BamHI sites of pET-19BHA.

p19BF1: Contains SREZBP zinc finger 1 (amino acids 204 to 241) fused in frame to the epitope tag in p19BHA. DNA encoding ZBP zinc finger 1 was generated by PCR using primers ZBPZF15' and ZBPZF13'. The resulting fragment was cut with XbaI and BamHI and ligated between the XbaI and BamHI sites of pET-19BHA.

p19BF123: Contains SREZBP zinc fingers 1, 2 and 3 (amino acids 204 to 297) fused in frame to the epitope tag in p19BHA. DNA encoding ZBP zinc fingers 1, 2 and 3 was generated by PCR using primers ZBPZF15' and ZBPZF33'. The resulting fragment was cut with XbaI and BamHI and ligated between the XbaI and BamHI sites of pET-19BHA.

Homeodomain Construct

p19BHH: Contains the Phox1 homeodomain and flanking amino acids (amino acids 43 to 150 (Grueneberg et al. 1992. Science. 257: 1089-1095)) fused in frame to the epitope tag in p19BHA. DNA encoding the Phox1 fragment was generated by PCR using primers Phox HH5' Primer and Phox HH Spe/Bam. The resulting fragment was cut with XbaI and BamHI and ligated between the XbaI and BamHI sites of pET-19BHA.

Zinc Finger/Homeodomain Constructs

p19B2FHH: Contains SREZBP zinc fingers 6 and 7 (amino acids 328 to 410) fused in frame to the epitope tag in p19BHA followed by the Phox1 homeodomain (amino acids 43 to 150). An XbaI-BamHI fragment from p19BHH

containing sequences encoding the Phox1 homeodomain was ligated between the SpeI and BamHI sites of p19B2F.

p19B4FHH: Contains SREZBP zinc fingers 4, 5, 6 and 7 (amino acids 300 to 410) fused in frame to the epitope tag in p19BHA followed by the Phox1 homeodomain (amino acids 43 to 150). An XbaI-BamHI fragment from p19BHH containing sequences encoding the Phox1 homeodomain was ligated between the SpeI and BamHI sites of p19B4F.

p19B7FHH: Contains SREZBP zinc fingers 1 to 7 (amino acids 216 to 410) fused in frame to the epitope tag in p19BHA followed by the Phox1 homeodomain (amino acids 43 to 150). An XbaI-BamHI fragment from p19BHH containing sequences encoding the Phox1 homeodomain was ligated between the SpeI and BamHI sites of p19B7F.

p19BZF1HH: Contains SREZBP zinc finger 1 (amino acids 204 to 241) fused in frame to the epitope tag in p19BHA followed by the Phox1 homeodomain (amino acids 43 to 150). An XbaI-BamHI fragment from p19BHH containing sequences encoding the Phox1 homeodomain was ligated between the SpeI and BamHI sites of p19BZF1.

p19BZF123HH: Contains SREZBP zinc finger 1, 2 and 3 (amino acids 204 to 297) fused in frame to the epitope tag in p19BHA followed by the Phox1 homeodomain (amino acids 43 to 150). An XbaI-BamHI fragment from p19BHH containing sequences encoding the Phox1 homeodomain was ligated between the SpeI and BamHI sites of p19BZF123.

Homeodomain/Zinc Finger constructs

p19BHH2F: Contains Phox1 homeodomain (amino acids 43 to 150) fused in frame to the epitope tag in p19BHA followed by ZBP zinc fingers 6 and 7 (amino acids 328 to 410). An XbaI-BamHI fragment from p19B2F containing sequences encoding ZBP zinc fingers 6 and 7 was ligated between the SpeI and BamHI sites of p19BHH.

p19BHH4F: Contains Phox1 homeodomain (amino acids 43 to 150) fused in frame to the epitope tag in p19BHA followed by ZBP zinc fingers 4, 5, 6 and 7 (amino acids 300 to 410). An XbaI-BamHI fragment from p19B4F containing sequences encoding ZBP zinc fingers 4, 5, 6 and 7 was ligated between the SpeI and BamHI sites of p19BHH.

p19BHH7F: Contains Phox1 homeodomain (amino acids 43 to 150) fused in frame to the epitope tag in p19BHA followed by ZBP zinc fingers 1 to 7 (amino acids 216 to 410). An XbaI-BamHI fragment from p19B7F containing sequences encoding ZBP zinc fingers 1 to 7 was ligated between the SpeI and BamHI sites of p19BHH.

p19BHHZF1: Contains Phox1 homeodomain (amino acids 43 to 150) fused in frame to the epitope tag in p19BHA followed by ZBP zinc finger 1 (amino acids 204 to 241). An XbaI-BamHI fragment from p19BZF1 containing sequences encoding ZBP zinc finger 1 was ligated between the SpeI and BamHI sites of p19BHH.

p19BHHZF123:Contains Phox1 homeodomain (amino acids 43 to 150) fused in frame to the epitope tag in p19BHA followed by ZBP zinc fingers 1, 2 and 3 (amino acids 204 to 297). An XbaI-BamHI fragment from p19BZF123 containing sequences encoding ZBP zinc fingers 1, 2 and 3 was ligated between the SpeI and BamHI sites of p19BHH.

Determination of the DNA binding specificity of Zinc finger/Homeodomain fusion proteins

Zinc finger/Homeodomain hybrid proteins were expressed using the Promega TnT coupled reticulocyte lysate system. 4 Micrograms of each of the following constructs was added to a 50 microlitre translation mix: p19B2FHH, p19B4FHH, p19B7FHH. p19BHH was also included as a positive control.

The DNA-binding specificity of Zinc Finger/Homeodomain hybrid proteins was determined as described (Pollock and Treisman, 1990, NAR. 18:6197-6204) except that 12CA5 antibody was used to immunoprecipitate protein-DNA

complexes. Four cycles of selection were performed, and the resulting fragments were amplified and cloned into pUC119 for analysis.

PCR Primers

SRE-ZBP

2F-Xba5': 5'-TCAGTCTAGATGTAACATATGCCAGAAAGCCTTC-3'

4F-Xba5': 5'-TCAGTCTAGATGCAAGGAGTGTGGAAAAACCTTT-3'

7F-Xba5': 5'-TCAGTCTAGATGTCATGAGTGTGGGAAAGCCTTT-3'

ZNF-Spe/Bam: 5'-TCAGGGATCCTCAATAACTAGTAGCCAGTTTGTCTTTGTGGTGATA-3'

ZBPZF15': 5'-TCAGTCTAGACATAAGAAAGTCCTCTCTAG-3'

ZBPZF13': 5'-TCAGGGATCCTCTATATCAACTAGTAGGCTTCTCACCAAGATGG-3'

ZBPZF33': 5'-TCAGGGATCCTCTATATCAACTAGTGGGCTCCTCCTGACTGTG-3'

PHOX1

Phox HH 5' Primer:
5'-TCAGTCTAGAGGCCGGAGCCTGCTGGAGT-3'

Phox HH Spe/Bam:
5'-TCAGGGATCCTCAATAACTAGTGTAGGATTTGAGGAGGGAA-3'

Claims:

1. A recombinant protein or protein complex comprising at least two DNA-binding subdomains, capable of binding to a selected or selectable DNA sequence.

2. A recombinant protein or protein complex of claim 1 which contains at least one zinc finger domain and at least one homeodomain.

3. A recombinant protein or protein complex of claim 1 in which the DNA-binding domain comprises components derived from human proteins.

4. A recombinant protein or protein complex of claim 1 which is capable of binding to a DNA sequence with a dissociation constant less than 10^{-9} M.

5. A recombinant protein or protein complex of claim 1 which further comprises a ligand binding domain.

6. A recombinant protein or protein complex of claim 5 in which the ligand binding domain comprises an FKBP domain.

7. A recombinant protein or protein complex of claim 1 which further comprises a transcriptional activating domain.

8. A recombinant protein or protein complex of claim 7 in which the transcriptional activating domain comprises a VP16 transcriptional activation domain.

9. One or more DNA sequences encoding a recombinant protein or protein components of a complex of any of claims 1-8.

10. An engineered cell containing and capable of expressing a DNA sequence or sequences of claim 9.

11. A DNA sequence comprising a target gene and a recombinant DNA sequence to which a recombinant protein or protein complex of any of claims 1-8 binds.

12. A method for expressing a target gene in cells which comprises providing cells of claim 10 which further contain a DNA sequence of claim 11 and maintaining them under conditions permitting gene expression.

13. A method for expressing a target gene in a cell which comprises
(a) introducing into cells (i) DNA encoding a recombinant protein or protein components of a complex of any of claims 1-8 and (ii) a second DNA sequence comprising a target gene and a DNA sequence to which the recombinant protein or protein complex binds; and
(b) maintaining the cells under conditions permitting continued cell growth and gene expression.

14. A method of claim 12 or 13 in which the conditions permitting continued cell growth and gene expression include maintaining the cells in a medium containing a suitable dimerizing agent in an amount effective to result in dimerization of the protein components of the protein complex capable of binding to the selected or selectable DNA sequence.

15. A kit comprising DNA encoding a recombinant protein or protein components of a complex of any of claims 1-8 and a second DNA sequence containing a target gene linked to a DNA element to which the recombinant protein or protein complex is capable of binding.

16. A kit comprising DNA encoding a recombinant protein or protein components of a complex of any of claims 1-8 and a second DNA sequence containing a cloning site linked to a DNA element to which the recombinant protein or protein complex is capable of binding.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/10557

A. CLASSIFICATION OF SUBJECT MATTER
IPC(6) :C07K 14/00; C12N 5/00, 15/00; C12P 21/06
US CL :530/350; 435/240.2, 69.1; 536/23.5
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 530/350; 435/240.2, 69.1; 536/23.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, DIALOG
search terms: DNA bind?, zinc finger?, homeodomain?, ligand?, FKBP, transcript?, activat?, VP16, chimer?, dimeriz?

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Mechanisms of Development, Volume 34, issued 1991, FORTINI ET AL, "The Drosophila zhf-1 and zhf-2 Genes Encode Novel Proteins Containing Both Zinc-finger and Homeodomain Motifs", pages 113-122, see entire document.	1-2, 4, 8-10
P - Y	Science, Volume 267, issued 06 January 1995, POMERANTZ ET AL, "Structure-Based Design of Transcription Factors", pages 93-96, see entire document.	1-4, 7-13, 15-16 --- 5, 14
Y	Cell, Volume 54, issued 15 July 1988, WEBSTER ET AL, "The Hormone-Binding Domains of the Estrogen and Glucocorticoid Receptors Contain an Inducible Transcription Activation Function", pages 199-207, see entire document.	5, 14

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search
OCTOBER 1995

Date of mailing of the international search report
20 NOV 1995

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
TERRY A. MCKELVEY
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/10557

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Science, Volume 262, issued 12 November 1993, SPENCER ET AL, "Controlling Signal Transduction with Synthetic Ligands", pages 1019-1024.	1-16
A	Science, Volume 251, issued 15 February 1991, WEINTRAUB ET AL, "The myoD Gene Family: Nodal Point During Specification of the Muscle Cell Lineage", pages 761-766, see page 372, Figure 1.	1-16
A, P	Molecular and Cellular Biology, Volume 15, Number 3, issued March 1995, QI ET AL, "The Ligand-Binding Domains of the Thyroid Hormone/Retinoid Receptor Gene Subfamily Function In Vivo To Mediate Heterodimerization, Gene Silencing, and Transactivation", pages 1817-1825.	1-16
A	Journal of Steroid Biochemistry and Molecular Biology, Volume 49, Number 2/3, issued 1994, BERGMANN ET AL, "Bivalent Ligands as Probes of Estrogen Receptor Action", pages 139-152.	1-16